DOCUMENT RESUME

ED 283 884 TM 870 424

AUTHOR Heller, Eric S.; Rife, Frank N.
TITLE Questionnaire Response Scales: Design Factors That Influence Respondent Satisfaction.
PUB DATE 23 Apr 87
NOTE 17p.; Paper presented at the Annual Meeting of the American Educational Research Association (Washington, DC, April 20-24, 1987).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Attitude Measures; Comparative Testing; Grade 7; Junior High Schools; *Participant Satisfaction; *Questionnaires; Rating Scales; *Response Style (Tests); School Surveys; *Test Construction; *Test Format
IDENTIFIERS Chi Square Analysis

ABSTRACT

The goal of this study was to assess the relative merit of various ranges and types of response scales in terms of respondent satisfaction and comfort and the nature of the elicited information in a population of seventh grade students. Three versions of an attitudinal questionnaire, each containing the same items but employing a different response scale, were administered randomly to three groups of seventh grade students. Specifically, a 2-point forced choice scale (agree-disagree), a traditional 5-point (strongly agree to strongly disagree) and a 21-point continuous scale which included a neutral position (strongly agree at zero to strongly disagree at 20) were compared. The results indicated the range of a response scale may not have much bearing on the seventh grade students' sense of comfort with the instrument. Providing neutral or undecided positions appears to increase rates of response and the reliability of the survey data. The range of a wide response scale does not seem to intimidate the seventh grade students, but they feel more comfortable when presented with specifically labelled points along that scale. (BAE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED283884

QUESTIONNAIRE RESPONSE SCALES: DESIGN FACTORS THAT INFLUENCE RESPONDENT SATISFACTION

By

Eric S. Heller
& Frank N. Rife

The Institute for Governmental Services
226 Middlesex House
University of Massachusetts
Amherst, MA 01003
(413)545-0001

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

E. Heller

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Paper presented at the annual meeting of the American Educational Research Association, April 23, 1987, in Washington, D.C.

TM 870 424

BEST COPY AVAILABLE

QUESTIONNAIRE RESPONSE SCALES: DESIGN FACTORS THAT INFLUENCE RESPONDENT SATISFACTION

OBJECTIVES & OVERVIEW

The measurement of human attitudes may be one of the most precarious of all tasks facing the social scientist. Knowledge can be normatively tested at least to some degree, and certain aspects of behavior can be observed, measured and documented. But what we can learn about the way people feel about an issue, their predispositions and internal emotional states is still to a large degree dependent on what they tell us, either in response to our direct questions or to our more ingenious probes. In as delicate a form of attitudinal measurement as a written questionnaire, variables which may possibly affect the nature of a person's response abound. The wording of questions, the length of the instrument, and its physical appearance are but a few of the factors typically cited and studied (Berdie, Anderson and Niebuhr, 1986; Sudman and Bradburn, 1982). It is also possible that the nature of the response options provided on such surveys may also have a bearing on the subjective frame of mind of a respondent and hence on the responses, particularly if a person does not feel that those options lend themselves to a clear expression of one's opinions.

The goal of the present study, therefore, was to assess the relative merit of various ranges and types of response scales in terms of respondent satisfaction and comfort and the nature of elicited information in a population of seventh grade students. Within an ongoing school evaluation study, different versions of an attitudinal questionnaire, each containing the same items but employing a unique response scale, were administered randomly to three groups of subjects. Specifically, two-point, five point, and twenty-one point scales were compared, and answers sought to the following methodological questions:

1. Do respondents feel more comfortable given limited scale points to choose from, or do they prefer more flexibility and range in their options?
2. What is the effect on respondents of a forced-choice format versus the inclusion of a neutral category of response?
3. Is the degree of respondent satisfaction with a given scale dependent upon degree of knowledge on and/or emotional involvement with the issues being surveyed?
4. Is the nature of elicited responses contingent upon respondent satisfaction and/or comfort with a particular response scale?

THEORETICAL BACKGROUND

No clear-cut guidelines exist for the design of response scales to Likert-type attitudinal survey items. Opinions differ throughout the professional literature with regard to both the optimal range and the inclusion or exclusion of options for neutral responses. Sheatsley (1983) suggests that beyond a five-point scale, problems arise with regard to scale design (the inability to attach words to each scale point), respondent ease of decision, and data analysis. Others, however, have prescribed various formats beyond the traditional five-point range (Green and Rao, 1970; Komorita and Graham, 1965). Sudman and Bradburn (1982) have recommended the inclusion of a middle, neutral category over forcing respondents to make an attitudinal commitment, while Sheatsley (1983) suggests that in certain circumstances such an option may lead respondents to avoid taking a difficult or possibly socially unpopular position.

Little evidence can be found of attempts to capture respondents' subjective perspectives on these methodological questions. Indeed, if respondent comfort and satisfaction have any bearing on the frequency of

incomplete surveys, overall response rates, or the reliability of data collected with such instruments, certainly more concrete data must be gathered on these issues.

Berdie (1986) raised questions of this nature through a post-hoc, indirect analysis of survey respondent satisfaction, and suggested the need for the systematic construction of research regarding how different populations both feel with and respond to various response scales. In light of the above, the present study was designed as a controlled experiment to provide comparative data on these issues for three unique response scale formats.

DATA COLLECTION

Instrumentation: As part of a larger evaluation of an instructional innovation, questionnaires were developed to survey students' perceptions and attitudes. Nine Likert statements were included in the instrument, and three variations of the questionnaire created, each employing one of the following response scales: 1) a two point forced choice scale (agree-disagree); 2) a traditional five-point range (strongly agree-agree-neutral-disagree-strongly disagree); and 3) a twenty-one point continuous scale which included a neutral position (strongly agree at 0 to strongly disagree at 20). Following the Likert items, a methodological question was presented to the respondents, reminding them of the type of scale they had utilized, and asking that they evaluate its appropriateness. Specifically, a respondent could indicate whether or not the responses provided always included one that matched her/his feelings on a given issue, and if not, what would have been preferable alternatives.

Respondents: The three versions of the instrument were systematically mixed and randomly distributed to 214 seventh grade students at a regional junior high school serving four communities. The students comprised the entire

population at that grade level, and were participating in an experimental program assessing a form of collaborative teaching. Complete randomization was maintained within and across all classes, and administration of the instrument was supervised by the researchers in the teachers' absence. Thus, all respondents responded to the same Likert statements, but each was offered only one of the three response formats.

ANALYSIS

Responses to the methodological question regarding satisfaction or dissatisfaction with the given response scale were compared among the three experimental conditions. Throughout this study, this measure of respondent satisfaction will be referred to as the 'a priori', subjective measure. An analysis of differential satisfaction levels among the three experimental conditions was tested by chi-square analysis.

Additionally, a post-hoc measure of dissatisfaction similar to Berdie's was applied as an additional means of testing the reliability of the subjective measure. Extraneous marks and comments on the instruments which clearly inferred dissatisfaction were tabulated and compared among the groups. Typical examples included exclamation marks next to an option, arrows connecting one option with another and written alternative responses. Incidence of one or more such signs of dissatisfaction on a form resulted in that subject being categorized as dissatisfied. As with the a priori measure, chi-square analysis was applied as a means of comparing the three groups on this measure.

For an analysis of the effect on respondent satisfaction of including or excluding a neutral mid-point in a response scale, the groups receiving the five-point and the twenty-one point scales were grouped as one (as each included a legitimate neutral option.) Dissatisfaction levels among respondents who were given those instruments, as determined both by the subjective and the

post hoc measures as above, were compared with those for respondents given the two-point, forced choice condition by means of the chi-square.

A qualitative review of the nature of dissatisfaction involved two sources of data -- content analysis (Borg and Gall, 1963) was performed on both the types of extraneous marks and comments which appeared on the scales themselves, as well as on the open-ended responses to the methodological question regarding preferable response options. (This question was addressed to those who subjectively reported dissatisfaction with the type of response option provided.) Through such an analysis, the marks, comments and responses were easily categorized and tabulated, as only a few, highly discrete modes of response arose.

An additional analysis was applied as a means of determining whether or not high emotional involvement and/or degree of familiarity of respondents with specific survey issues results in a need for more discrimination in a response scale. It has been suggested that people with more knowledge or interest in a survey topic require the largest number of response options (Berdie, 1986). Based on this observation, we hypothesized that people would be more discriminating on issues most directly affecting them than they would be on other issues of lesser interest to them.

To this end, at the time of instrument development the nine items on the questionnaire were grouped according to the level of emotional interest each was expected to hold for the target population. Pre-questionnaire interviews had revealed that certain issues surrounding the collaborative teaching experiment, particularly those with social implications for the students, provoked highly charged and extreme responses from the students, as they apparently addressed more urgent needs and concerns than did other items. The items were grouped in three subsets reflecting high, medium and low expected

levels of emotional content. Employing the post hoc measure of dissatisfaction (the subjective measure was inappropriate for this task as it addressed the instrument as a whole and not the individual items), comparisons of the three experimental groups were carried out for each subset of questions.

The evaluation of the consistency of the responses collected with the different instruments and any effects of respondent satisfaction on the breakdown of those responses was performed in two stages. First, a straight comparison of responses between the five and twenty-one point scales was carried out (with chi-square analysis) for each of the nine items on the questionnaire. For this analysis, responses were recoded into three categories: agree, neutral, and disagree. Clearly, the two-point scale did not lend itself to such a comparison, as no neutral responses were recognized.

An attempt to assess any effects of the exclusion of the neutral option on the breakdown of responses was slightly more complex. The assumption was made that if this factor had no significant effect, then given the opportunity, equivalent numbers of those who would otherwise have agreed and those who would have disagreed would choose the neutral option. To test this assumption, the five and twenty-one point scales were grouped as one, and for each item, only those cases in which 'neutral' was not chosen were selected, and compared with responses to the two-point scale. Were significant differences in the response patterns to arise, we would have to conclude that the assumption is not valid.

RESULTS

A. COMPARISON OF THE THREE SCALES

Table 1 summarizes the relevant data and presents the results of the analysis of differences in satisfaction levels among the three experimental conditions. While the level of dissatisfaction is higher for the two-point scale respondents, the similarity between the levels for the other scales

results in a non-significant statistical difference, regardless of the measure employed. Under none of the conditions are the levels of dissatisfaction negligible, however, and the overall levels of dissatisfaction are 32% subjectively and 24% through the post hoc analysis. Also of interest is the fact that the post hoc measure is consistently lower than the subjective report. Possibly, in some cases the prompt itself caused reflection and resultant dissatisfaction for respondents who were otherwise not dissatisfied enough to express this state on the response scales.

TABLE 1: Relative Rates of Respondent Dissatisfaction With the Three Response Scale Options

	A Priori (Subjective) Measure	Post Hoc Measure of Extraneous Marks
Response Scale Type	% Dissatisfied (N=192)*	% Dissatisfied (N=214)
2-Point Forced Choice (N=72)	42%	32%
5-Point Discrete (N=71)	27%	22%
21-Point Continuous Scale (N=72)	28%	17%
Overall Dissatisfaction	32%	24%
Chi-square value	4.314	4.849
Degrees of Freedom	2	2
Significance level P	0.1156	0.0885

* 22 respondents failed to respond to the subjective probe.

B. INCLUSION/EXCLUSION OF A NEUTRAL RESPONSE

Data comparing levels of satisfaction between the two-point, forced choice scale and the combined five and twenty-one point scales are summarized in Table 2. Significant differences in the levels of dissatisfaction under the two conditions are reflected both with the subjective and post-hoc measures. A higher percentage of those denied the opportunity to express a neutral position reported subjectively and exhibited explicitly some degree of dissatisfaction with the response options. Whether or not a causal relationship can be interpreted here rests considerably on the nature of the respondents' dissatisfaction. [See Section C]

TABLE 2: Comparison of Levels of Dissatisfaction for Scales With and Without a Neutral Response Option

Response Scale Type	A Priori (Subjective) Measure % Dissatisfied (N=192)*	Post Hoc Measure of Extraneous Marks % Dissatisfied (N=214)
Neutral Response Option Provided (N=143)	27%	20%
Forced Choice Only (No Neutral Option) (N=71)	42%	32%
Overall Dissatisfaction	32%	24%
Chi-square value	4.230	4.292
Degrees of Freedom	1	1
Significance level P	0.0381**	0.0383**

* 22 respondents failed to respond to the subjective probe.

** $P < 0.05$

C. THE NATURE OF DISSATISFACTION

Content analysis of the types of responses to the methodological question regarding preferable response options and of the nature of the extraneous marks and comments on the body of the instrument revealed a number of discrete, clearly defined causes of dissatisfaction. Specifically, the following categories were determined:

- Need for a neutral response category
- Need for more intermediate response categories along the scale
- Need for more extreme labels for the scale poles
- Need for opportunity to elaborate opinions (open-ended responses)

Tables 3 and 4 summarize the frequency of occurrence of each category of response within each of the experimental groups and over all respondents for the two sources of data (a priori and post hoc):

TABLE 3: The Nature of Dissatisfaction With Response Scales Measured by Responses to the Methodological Question

	Response Scale Type			
Type of Dissatisfaction	2-Point % (n)	5-Point % (n)	21-Point % (n)	Overall % (n)
Neutral Category Desired	55% (11)	—	—	31% (11)
More Intermediate Categories Desired	25% (5)	63% (5)	57% (4)	40% (14)
More Extreme Labels for Poles Desired	20% (4)	—	—	12% (4)
Open-ended Response Option Desired	—	37% (3)	43% (3)	17% (6)

TABLE 4: The Nature of Dissatisfaction With Response Scales Measured by the Analysis of Extraneous Marks and Written Comments

Type of Dissatisfaction	Response Scale Type: 2-Point % (n)	5-Point % (n)	21-Point % (n)	Overall % (n)
Neutral Category Desired	56% (15)	—	—	27% (15)
More Intermediate Categories Desired	11% (3)	63% (10)	33% (4)	31% (17)
More Extreme Labels for Poles Desired	33% (9)	37% (6)	67% (8)	42% (23)
Open-ended Response Option Desired	—	—	—	—

* Note that in Tables 3 and 4, the percentages are of the total number of responses to the question or marks of dissatisfaction.

Clearly, the overwhelming cause of dissatisfaction for those receiving the two-point scale was the lack of a neutral mid-point. It would also appear that in no case was a particular scale viewed as being too long or detailed. Rather, if anything, respondents preferred more categories of response and precise definitions of points along the scale.

D. EMOTIONAL INVOLVEMENT OF RESPONDENTS

To measure the impact of the emotional content of items on respondents' satisfaction with the scales, the nine questions were grouped as high, medium or low according to earlier group interview responses. Table 5 summarizes the results of the comparison of satisfaction levels within each of these subsets of items, as determined by the post hoc measure.

TABLE 5: Respondent Satisfaction With the Three Scales as a Function of the Emotional Content of the Items

	Emotional Content of Items		
	High	Medium	Low
Response Scale Type	% Dissatisfied	% Dissatisfied	% Dissatisfied
2-Point Forced Choice	17%	7%	21%
5-Point Discrete	10%	8%	11%
21-Point Continuous Range	7%	4%	7%
Chi-square value	3.708	1.030	6.564
Degrees of Freedom	2	2	2
Significance level P	0.1566	0.5974	0.0376*

* $P < 0.05$ level.

As the data in Table 5 reveal, the only category of questionnaire items for which significant differences in satisfaction can be detected is that of the expected low emotional content. The results of this analysis are clearly contrary to the earlier contention that as the respondents' emotional involvement increases, their need for more discriminating categories of response also grows. Two possible explanations are offered. If indeed the original hypothesis is true, then quite possibly the subjective grouping of the items on the questionnaire was misinformed with regard to the respondents' knowledge of the issues and resultant emotional involvement. If, on the other hand, the groupings were accurate, then an alternative hypothesis would be that the less emotionally involved respondents are with an issue, the more objective they will be and therefore more discriminating in their choice of response. Certainly both interpretations warrant further investigation in subsequent studies of this nature.

E. Nature of the Responses

The initial evaluation of the consistency of data collected via instruments employing different response scales involved the comparison of responses to the nine items by those using the five point scale and those using the twenty-one point scale, as both scales included neutral options. Analyses were performed individually on each of the nine items. Consistently, no significant differences in response patterns were discerned for any of the questions. It can be concluded, therefore, that the nature of responses to these questions was not differentially effected by either of these scales. Table 6 presents the results of the chi-square analyses for the individual questionnaire items:

Table 6: Comparison of Responses to the 5-Point and 21-Point Scales

Item #	Chi-Square	Degrees of Freedom	Significance P
1	3.404	2	.182
2	0.756	2	.685
3	1.164	2	.559
4	2.329	2	.312
5	2.733	2	.255
6	2.714	2	.257
7	0.991	2	.609
8	0.239	2	.888
9	0.330	2	.848

As described previously [see Analysis], the comparison of responses to the two point scale was made with those to the other scales who had not chosen the

neutral response on a given question. Here the results were less consistent than in the previous analysis. For three of the nine items, response patterns were significantly different for these two conditions. (Interestingly, these were the three questions categorized as having medium emotional importance to the respondents.) It would appear that for these items, it is incorrect to assume that the inclusion of a neutral point has no bearing on response trends. Rather, the neutral option has significantly altered the nature of the findings. What direction that effect has is unique to each question, and most probably a function of the nature of the issues with which each item deals.

The results of these analyses are presented in Table 7:

Table 7: Comparison of Responses to the 2-Point and Combined 5 and 21-Point Scales (Excluding Neutral Responses)

Item #	Chi-Square	Degrees of Freedom	Significance P
1	4.220	1	.040*
2	0.206	1	.650
3	13.352	1	.000*
4	0.287	1	.592
5	1.148	1	.284
6	1.990	1	.158
7	5.529	1	.019*
8	0.001	1	1.000
9	2.328	1	.127

* $P < 0.05$.

CONCLUSIONS

While it appears that the range of a response scale may not have critical bearing on seventh graders' sense of comfort with an instrument, legitimizing the expression of neutral or undecided positions on survey issues may indeed serve to benefit their subjective frame of mind and hence increase rates of response and the reliability of survey data. In some cases, it will often alter the resulting picture captured by the survey instrument. Typically, it seems that this population is more frustrated when presented general response categories that do not necessarily capture their opinions closely enough. While the range of a wide response scale does not seem to intimidate such respondents, they feel more comfortable when presented with specifically labelled points along that scale. Clearly we are limited by our language in the degree of discrimination for which we can provide meaningful scale labels.

A deeper understanding of the effects of such variables on respondent satisfaction and hence, clearer guidelines for the tailored design of questionnaire response scales will certainly begin to evolve as future research investigates varying types of response scales on differing types of respondent populations.

REFERENCES

Berdie, D.R. (1986). "The optimum number of survey research scale points: What respondents say." Paper presented at the meeting of the American Educational Research Association, San Francisco, CA.

Berdie, D.R., Anderson, J.F. & Niebuhr, M.A. (1986). Questionnaires: Design and Use. Metuchen, N.J.: Scarecrow Press, Inc.

Green, P.E. & Rao, V.R. (1970). "Rating scales and information recovery: How many scales and response categories to use?" Journal of Marketing, 34, pp. 33-39.

Komorita, S.S. & Graham, W. (1965). "Number of scale points and the reliability of scales," Educational and Psychological Measurement, 25, pp. 987-995.

Sheatsley, P.B. (1983). "Questionnaire construction and item writing," in Rossi, P.H., Wright, J.D. & Anderson, A.B. (eds.) Handbook of Survey Research. Orlando: Academic Press, Inc.

Sudman, S. & Bradburn, N.M. (1982). Asking Questions: A Practical Guide to Questionnaire Design. San Francisco: Jossey-Bass.